UNCLASSIFIED MASTER COPY - FOR REPRODUCTION PUBLISHERS (2)

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

AD-A213 869

1a.	1b. RESTRICTIVE MARKINGS
2a.	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b.	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARO 25333.5-MA-SDI
6a. NAME OF PERFORMING ORGANIZATION Utah State Univ	6b. OFFICE SYMBOL (If applicable)
7a. NAME OF MONITORING ORGANIZATION U. S. Army Research Office	
6c. ADDRESS (City, State, and ZIP Code) Logan, UT 84322	7b. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U. S. Army Research Office	8b. OFFICE SYMBOL (If applicable)
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAAL-03-88-K-0051	
8c. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Fast Algorithms for Estimating Mixture Parameters

12. PERSONAL AUTHOR(S)

Homer F. Walker

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 7/1/88 TO 6/30/89	August 30, 1989	5

16. SUPPLEMENTARY NOTATION

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Fast Algorithms, Algorithms, Mixture Distribution Random Variables

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The investigation is a two year project with the first year sponsored by the Army Research Office and the second year by the National Science Foundation (Grant Number: DMS-0088995). The results reported are those obtained under the sponsorship of the Army Research Office during the period July 1, 1988 to June 30, 1989. During this period the work on the accelerated scoring technique was completed and prepared for publication, and the numerical testing of the accelerated fixed-point method was completed. The work on relaxation methods will be done under the sponsorship of the National Science Foundation during the coming year.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

89 10 24 216

ARO 25333.5-MA-SDI

FAST ALGORITHMS FOR ESTIMATING MIXTURE PARAMETERS

FINAL REPORT

Homer F. Walker and Michael P. Windham

August 30, 1989

U. S. ARMY RESEARCH OFFICE

Grant Number: DAAL03-88-K-0051

UTAH STATE UNIVERSITY

A-i

Fast Algorithms for Estimating Mixture Parameters[1]

A. Problem:

The project consists of investigating three numerical methods estimating the parameters of a mixture distribution. The estimates for the parameter θ are obtained by maximizing the log-likelihood function

$$L(\theta) = \sum_k p(x_k|\theta) = \sum_k p_k$$

where $\{x_1, \ldots, x_n\}$ is a sample of size n from a mixture distribution.

1. Accelerated Scoring: The first approach depends on the fact that the log-likelihood function for a mixture distribution has the form

   $$L''(\theta) = -JJ^T + \sum_k p_k''/p_k.$$

   The first term of L'' involves only first derivatives and is easy to compute relative to the second term. Therefore, the bulk of the computation in using Newton's method to estimate θ would involve obtaining the second derivative term. This phase of the project involved replacing the second derivative term with various quasi-Newton approximations.

2. Accelerated Fixed-point: The EM algorithm is often used for obtaining parameter estimates for mixtures. It can be viewed as a procedure for finding a solution to an equation of the form $\theta = G(\theta)$. Using Newton's method to solve this problem requires the computation of $I - G'(\theta)$. This phase of the project investigates replacing the derivative by a quasi-Newton approximation obtained by computing updating vectors based on several EM steps and applying a quasi-Newton update to accelerate the convergence.

3. EM Relaxation: This phase of the project is investigating the effect of a relaxation of the form

   $$\theta_{N+1} = \frac{1}{1-\lambda}\theta_N - \frac{\lambda}{1-\lambda}G(\theta_N)$$

[1]The view, opinions, and/or findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision unless so designated by other documentation.

where

$$\lambda = \frac{(\theta_N - \theta_{N-1})^T(\theta_{N-1} - \theta_{N-2})}{|\theta_{N-1} - \theta_{N-2}|^2}.$$

The parameter λ is an approximation to the largest eigenvalue of G' and so this approach is a generalization of the one-dimensional Aitken acceleration.

B. Results:

The investigation is a two year project with the first year sponsored by the Army Research Office and the second year by the National Science Foundation (Grant Number: DMS-0088995). The results reported are those obtained under the sponsorship of the Army Research Office during the period July 1. 1988 to June 30, 1989. During this period the work on the accelerated scoring technique was completed and prepared for publication, and the numerical testing of the accelerated fixed-point method was completed. The work on relaxation methods will be done under the sponsorship of the National Science Foundation during the coming year.

Accelerated Scoring: Newton-like iterative procedures for maximizing L have the form

$$\theta_{N+1} = \theta_N - B_N^{-1}\nabla_\theta L(\theta_N)$$

where the method is determined by the choice of B_N. Newton's method, itself, uses $B_N = L''(\theta_N)$. As was mentioned above $L''(\theta) = C(\theta) + A(\theta)$, where $C = -(\sum_k \nabla_\theta p_k/p_k)(\sum_k \nabla_\theta p_k/p_k)^T$ and $A = \sum_k p_k''/p_k$. The Method of Scoring (MOS) uses updates with $B_N = C(\theta_N)$, which simply ignores the second derivative term under the supposition that in general its effect is small.

We have developed a method, called *MLE updating*, which does not ignore the second derivative but uses $C(\theta_N)$ together with a quasi-Newton approximation to $A(\theta_N)$. The update is given by

$$A_{N+1} = A_N + \frac{w_N s_N^T + s_N w_N^T}{v_N^T s_N} - \frac{s_N^T w_N}{(v_N^T s_N)^2} v_N v_N^T$$

where

$$\begin{aligned}
s_N &= \theta_{N+1} - \theta_N \\
v_N &= \nabla_\theta L(\theta_{N+1}) - \nabla_\theta L(\theta_N) \\
w_N &= \sum_k (\nabla_\theta p(x_k|\theta_{N+1}) - \nabla_\theta p(x_k|\theta_N))/p(x_k|\theta_{N+1}) - A_N s_N.
\end{aligned}$$

Theoretical Results

THEOREM: *If $\hat{\theta}$ is a point for which $\nabla_\theta L(\hat{\theta}) = 0$, $\nabla_\theta L$ is differentiable on an open convex neighborhood Ω of $\hat{\theta}$, C is continuous at $\hat{\theta}$, there are $\gamma \geq 0$ and $p \in (0, 1]$ such that*

$$|L''(\theta) - L''(\hat{\theta})| \leq \gamma |\theta - \hat{\theta}|^p$$

for all $\theta \in \Omega$, $L''(\hat{\theta})$ is invertible, there is a $\gamma_C \geq 0$ so that

$$|C(\theta) - C(\hat{\theta})| \leq \gamma_C |\theta - \hat{\theta}|^p,$$

then there are $\epsilon > 0$ and $\delta > 0$ such that if $|\theta_0 - \hat{\theta}| < \epsilon$ and $|A_0 - A(\hat{\theta})| < \delta$, the MLE updates are well-defined and converge q-superlinearly to $\hat{\theta}$. Furthermore, $\{|B_N|\}_{N=0,1,\ldots}$ and $\{|B_N^{-1}|\}_{N=0,1,\ldots}$ are uniformly bounded.

This theorem shows that the convergence properties of the MLE updating method are comparable to the well-known least squares methods.

Numerical Experiments

Numerical experiments where conducted with samples from mixtures of univariate normals. The parameters used are detailed in Gonglewski and Walker (see **C.** below), but here we summarize in Table 1. representative results obtained using five methods

1. MLE updating (MLE)
2. Finite Difference Hessian (FDH)
3. Method of Scoring (MOS)
4. Broyden-Fletcher-Goldfarb-Shanno Updating (BFGS)
5. EM Algorithm (EM)

for samples with various distances between population means, $\Delta\mu$.

This testing shows that MLE updating compares very favorably with standard general optimization methods. Both the MLE and standard general optimization methods seem likely to outperform the EM algorithm on mixture estimation problems, especially when the component populations are poorly separated.

Table 1.

$\Delta\mu$	MLE	FDH	MOS	BFGS	EM
6.0	13	11	14	24	5
4.0	13	7	13	22	44
2.0	10	8	15	26	883
1.0	22	23	27	46	777
0.4	42	37	38	76	1381
0.2	54	51	59	87	3095

Iterations required to obtain $|\theta_N - \hat{\theta}|_\infty \leq 10^{-6}$

Accelerated Fixed-point: A Newton acceleration of the EM algorithm could be accomplished by periodically inserting between EM iterations, $\theta_{N+1} = G(\theta_N)$, an update of the form

$$\theta_{N+1} = \theta_N - F'(\theta_N)^{-1}F(\theta_N)$$

where $F(\theta) = \theta - G(\theta)$. Our quasi-Newton method replaces the inverse of the derivative by a Broyden-update approximation

$$B_{N+1}^{-1} = (I + v_N s_N^T)B_N^{-1}$$

where $s_N = G(\theta_N) - \theta_N$ and

$$v_N = \frac{s_N - B_N^{-1}(F(G(\theta_N)) - F(\theta_N))}{s_N^T B_N^{-1}(F(G(\theta_N)) - F(\theta_N))}.$$

This approach does not require the computation of a derivative and only the vectors s_N and v_N need be stored.

The graduate students Wick and Shea, under the direction of Walker and Windham, have constructed an experimental code which, according to options selected by the user, generates samples on mixtures of univariate or multivariate normally distributed random variables and determines approximate maximum-likelihood estimates using either the unmodified EM algorithm or the EM algorithm accelerated via Broyden updating. Preliminary computational results look promising, but a good bit of additional coding and numerical experimentation needs to be done to bring this phase of the research to completion. The preliminary results show that one can expect that the number of iterations required for well-separated populations is the same with or without updating, but when the populations are poorly separated updating reduces the computation by fifty percent in general and as much as seventy percent.

The work on the EM algorithm for mixtures has also led to a very promising line of investigation bearing on perhaps the most difficult problem associated with mixture estimation, determining the number of component populations in a mixture. This investigation centers around a certain "information ratio" matrix involving the Fisher information matrices of the mixture data and of the "labeled" data given the mixture data. Windham, with the collaboration of Walker and H.-H. Bock of the Rheinisch-Westfällische Technische Hochschule, Aachen, FRG, has shown that the eigenvalues of this "information ratio" matrix, especially the extreme eigenvalues, are closely related both to the number of component populations in a mixture and the speed of convergence of the EM algorithm. Thus the observed speed of convergence of the EM algorithm may be used to evaluate the validity of assumptions about the number of component populations in the mixture.

C. Acknowledging publications:

J. D. Gonglewski and H. F. Walker, *Quasi-Newton methods for maximum-likelihood estimation*, submitted for publication in Proc. 5^{th} Numerical Analysis Workshop, S. Gómez, J.-P. Hennart, and R. A. Tapia, eds., to be published by SIAM; also Utah State University Math. Stat. Dept. Res. Report August/89/47.

M. P. Windham, H. H. Bock and H. F. Walker, *Cluster Validity from Information Ratios*, Utah State University Math. Stat. Dept. Res. Report July/1989/46.

M. P. Windham and H. F. Walker, *Fisher Information from convergence rates of the EM algorithm*, Utah State University Math. Stat. Dept. Res. Report April/1989/43.

The following publications were completed during the period of Army Research Office sponsorship and acknowledged the support. They resulted from work in the general area and were often motivated by the work on this project. The results and their relation to this project appear in the Progress Reports.

S. K. Bourji and H. F. Walker, *Least-change secant updates of nonsquare matrices*, in revision for SIAM J. Numer. Anal.; also Utah State Uni-

versity Math. Dept. Res. Report, December 1987/38, Revision 1, May 1989.

H. F. Walker, *Implementations of the GMRES method*, invited paper in *Proceedings of the Workshop on Practical Iterative Methods for Large Scale Computations*, held at Minnesota Supercomputer Center, Minneapolis, Minnesota, October, 1988, published in Computer Physics Communications, 53 (1989), pp. 311-320; also Utah State University Math. Stat. Dept. Res. Report August/88/40 and Lawrence Livermore Nat. Lab. Report.

H. F. Walker, *Newton-like methods for underdetermined systems*, invited paper in *Proc. AMS-SIAM Summer Seminar on Computational Solution of Nonlinear Systems of Equations*, published in AMS *Lectures in Applied Mathematics* (to appear); also Utah State University Math. Stat. Dept. Res. Report November/88/42.

H. F. Walker and L. T. Watson, *Least-change secant update methods for underdetermined systems*, in revision for SIAM J. Numer. Anal.; also Utah State University Math. Stat. Dept. Res. Report August/88/41, VPI&SU Comp. Sci. Dept. Tech. Rep. 88-28, and VPI&SU Interdisciplinary Center for Applied Math. Rep. 88-09-03, September, 1988, Revision 1, June, 1989.

M. P. Windham, *Statistical models in cluster analysis*, Utah State University Math. Stat. Dept. Res. Report May/1989/45.

M. P. Windham, H. H. Bock and H. F. Walker, *Cluster Validity from Information Ratios*, Utah State University Math. Stat. Dept. Res. Report July/1989/46.

M. P. Windham and H. F. Walker, *Fisher Information from convergence rates of the EM algorithm*, Utah State University Math. Stat. Dept. Res. Report April/1989/43.

D. Personnel:

Senior personnel: Michael P. Windham (July - August, 1988); Homer F. Walker ($2\frac{1}{4}$ months, October, 1988 - June, 1989).

Graduate students: Darren Wick (July - August, 1988, Master of Science degree completed August 1989); Gary Shea ($5\frac{1}{2}$ months, October, 1988 - March, 1989).